DOCUMENT RESUME

ED 453 167 SP 039 964

AUTHOR Allen, James D.; Lambating, Julita
TITLE Validity and Reliability in Assessment and Grading: Perspectives of Preservice and Inservice Teachers and Teacher Education Professors.
PUB DATE 2001-04-00
NOTE 34p.; Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001). Figures may not reproduce adequately.
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *Grading; Higher Education; Preservice Teacher Education; *Reliability; *Student Evaluation; *Student Teacher Attitudes; Student Teachers; Teacher Educators; *Validity

ABSTRACT

This study investigated preservice, inservice, and college teachers' reasons for making assessment and grading decisions. A survey examined their perspectives about a case study that presented a grading dilemma in which a teacher, "Sarah", assigned grades based on criteria that were potentially invalid and unreliable. Participants analyzed the case and provided solutions and a rationale for their decisions. The dilemma was that Sarah had given the most knowledgeable student in her math class, "James", a grade of B because he never turned in homework. James' parents wanted Sarah to change James' grade to an A to reflect his mathematics knowledge. Participants discussed: whether Sarah should change James' grade, what grade James deserved, and what they believed was the purpose of homework and grades. Most respondents believed James' grade should not be changed because of the teacher's homework requirement. Results highlight the need to help teachers make good grading decisions based on fundamental measurement principles. In situations in which a teacher must summarize and communicate a student's classroom progress in an academic subject through one grade, there must be consensus that the grade represents the most accurate statement of the student's academic achievement, and only academic achievement. (Contains 57 references and 4 figures.) (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 453 167

Validity and reliability in assessment and grading: Perspectives of preservice and inservice teachers and teacher education professors

James D. Allen, Ph.D.
The College of Saint Rose
Albany, New York

Julita Lambating, Ph.D.
The California State University
Sacramento, California

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

J. D. Allen

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, April 2001
Session 21.59

SP039964

BEST COPY AVAILABLE

Introduction

Teachers have the responsibility to assess and grade students' academic achievement. Not only is grading a major responsibility of classroom teachers it is also a practice with which they, as well as preservice teachers, are often uncomfortable and find most difficult (Barnes, 1985; Lomax, 1996; Thorndike, 1997). The sources of this discomfort and difficulty seem to be three-fold. First there is the question of what student activities should constitute "academic achievement" and how to handle ancillary features of achievement such as a students' efforts. Although ancillary information such as effort and attitude could be part of an overall student report, they should not be part of a grade that should represent academic achievement (Tombari & Borich, 1999). Second, teachers often seem to be unsettled regarding the communication function of grades and they often try to communicate ancillary information about students that is not possible with a single academic grade. A third difficulty is in a large part due to the fact that many teachers make grading decisions that are not built on a solid foundation of relevant principles of measurement as it relates to assigning meaningful grades (Cizek, 1996; Marzano, 2000). Two of the most fundamental measurement principles related to assessment and grading are the principles of validity and reliability (Gallagher, 1998; Gredler, 1999; Linn & Gronlund, 2000; Stiggins, 2001).

Although there are many validity and reliability issues involved in classroom assessment, such as the variety and types of assessment instruments constructed and used by teachers and the types of learning behaviors being assessed (Ormrod, 2000), the focus of this paper is on the valid and reliable communication of final class grades as summaries of students' academic achievement. This research explores the perspectives of four groups of educators on the assessment and grading of students. It attempts to understand how these educators believe grades

should be used in summarizing and communicating students' performance and to what level they consider the fundamental principles of validity and reliability in making these judgements.

As illustrated by the title of the 1996 Yearbook of the Association for Supervision and Curriculum Development, *Communicating Student Learning* is an important function of schools and teachers (Guskey, 1996). Although there are various means to communicate student learning, currently a single report card grade for each academic subject is the most common and generally accepted system in upper elementary and secondary schools (Bailey & McTighe, 1996; Lake & Kafka, 1996). Bailey & McTighe (1996) argue that as a communication system, educators need to address four elements in their grading practices: the content of what grades communicate, the process of how grades are communicated, the purpose of why grades are being communicated, and to whom the grades are being communicated. They sum up these four elements in stating that "the primary purpose of secondary level grades and reports [is] to communicate student achievement to students, parents, school administrators, postsecondary institutions, and employers" so that each can make accurately informed decisions about the student's future (p. 120). The larger the variability in grading practices from teacher to teacher and from school to school, the more limited the value grades have as guides for planning the academic and career futures of students (Thorndike, 1997).

Validity, as used in this paper, addresses the accuracy of the assessment and grading procedures used by teachers (Gallagher, 1998; Gredler, 1999; Linn & Gronlund, 2000). Most fundamentally, does the assessment procedures and assignment of grades accurately reflect and communicate the academic achievement of the student? Reliability considers the dependability of the assessment and grading (Gallagher, 1998; Gredler, 1999; Linn & Gronlund, 2000). Does the assessment and grade provide a dependable measure and communication of what the student

has learned? The issue of validity is particularly important because, as we argue in this paper and supported elsewhere, the sole purpose of grades is to accurately communicate to others the level of academic achievement that a student has obtained (Snowman, Biehler & Bonk, 2000). If the grades are not accurate measures of the student's achievement, then they do not communicate the truth about the student's academic achievement or level of mastery of a subject. So even as "grades continue to be relied upon to communicate important information about performance and progress ... they probably don't" (Cizek, 1996, p. 104). Since important decisions are often based on a student's grade, unreliable and invalid grades may result in dire consequences for the student. Invalid grades that communicate an understatement of the student's understanding may prevent a student with ability to pursue certain educational or career opportunities. Grades that communicate an overstatement of the student's understanding of a subject may place a student in situations for which he or she is not prepared (Stumpo, 1997). This may be especially a problem with some teachers' grading practices in schools in poverty areas. Research indicates that when compared to schools in more affluent areas, students in low SES schools receive grades that are two letter grades better than students in affluent schools when national standardized scores are held constant (National Education Longitudinal Study, 1988). That is, students from schools in poor communities are receiving grades of B and C, while students from affluent schools, who have the same standardized scores, are receiving grades of D and F. Even if one assumes that there may be some bias in standardized tests, it seems clear that some grade inflation is likely occurring in schools with a large number of lower SES students. This may account for why students entering college from schools located in economically poor districts have a lower graduation rate than those from more affluent districts.

One of the goals of a teacher education program should be to prepare preservice and inservice teachers to develop effective methods to assess students and to communicate clearly and accurately through their grading practices that assessment to others. These methods and practices should be based on sound principles of measurement theory. It is the hypothesis of this study that as we enter a new millennium many preservice and inservice teachers, as well as teacher-educators, base many assessment and grading judgements not on sound principles of validity and reliability, but on other personally experienced, non-theory based, value-laden criteria. These criteria based on personal educational experiences and values can decrease the validity and reliability of the assessments and the grades given to students by teachers (Stiggins, 2001).

Review of the literature

Cizek (1996) argues that the "lack of knowledge and interest in grading translates into a serious information breakdown in education" and that "reforming classroom assessment and grading practices will require educators' commitment to professional development, classroom-relevant training programs" (pg. 103). Cizek's statement implies at least two aspects of grading that need to be addressed: (1) the role of grades in communicating accurate information about students to various interested parties, and (2) the training of teachers in grading practices based on sound measurement principles relevant to their classroom lives.

The purpose of grades and criteria for assigning grades is a topic in major texts devoted to classroom assessment (Airasian, 2000; Gallagher, 1998; Gredler, 1999; Linn & Gronlund, 2000; Nitko, 2001; Oosterhof, 2001; Stiggins, 2001). All these textbook authors suggest that the major reason for assigning grades is to create a public record of students academic achievement which can accurately and effectively communicate to others the level of mastery of a subject a

student has demonstrated. Nitko (2001) points out that: "Grades ... are used by students, parents, other teachers, guidance counselors, school officials, postsecondary educational institutions, and employers. Therefore [teachers] must assign grades with utmost care and maintain their validity" (p. 365). However, according to Marzano (2000), in contrast to teachers', students', parents', and community members' assumption that "grades are reliable measures of student achievement ... grades are so imprecise that they are almost meaningless" (p. 1). Thus, if a single grade on a report card or transcript is to effectively communicate information to all these varied parties, that single grade has to have some shared and accurate meaning (O'Connor, 1995). Friedman & Frisbie (1995; 2000) make a particularly strong argument for making sure that report card grades validly report information to parents about a student's academic progress and that teachers and administrators share a common understanding of what information a grade should communicate. This is why many argue that the only way to insure clear and valid (accurate) communication about students' progress in school is to make sure that the grade only includes criteria that measure students' academic achievement (Friedman & Frisbie, 2000; Ornstein, 1994). Including other factors, such as students' effort, motivation, discipline, or attitude, complicates the ability to interpreted a grade since these factors may be in direct conflict to each other and distort the meaning of a grade (Cross & Frary, 1996; Guskey, 1994; Linn & Gronlund, 2000; Nitko, 2001; Stiggins, 2001, Stumpo, 1997).

Grading practices by teachers rarely follow the measurement principles and grading practices recommended in measurement textbooks (Cross & Frary, 1996; Frary, Cross & Weber, 1993). Grading systems used by teachers vary widely and unpredictably and often have low levels of validity and reliability due to the inclusion of non-academic criteria used in the calculation of grades (Brookhart, 1994; Cizek et al., 1995; Cross & Frary, 1996; Frary et al.,

1993; Friedman & Frisbie, 1995; Johnsen, 1995; Olson, 1989). Some believe that this is due to the lack of adequate training of teachers in objective and systematic methods of evaluation leaving teachers to rely on poor grading practices (Stiggins, 1999; 1988). Others argue that even when teachers are provided with measurement instruction, they still use subjective value judgements when assigning grades (Brookhart, 1993).

A common finding of research on classroom teachers and their grading practices indicate they are often confused or unclear and often contradictory in their use of which criteria should be used to assign grades. Non-academic factors are often used as criteria for assigning grades because some teachers consider the consequences of grades over the value of clear communication of information and interpretability of the grades (Brookhart, 1993). Teacher have been found to make decisions about grades related to student effort in attempts to be "fair" in their grading practices (Barnes, 1985). Some studies have found that 2 out of 3 teachers believe that effort and student conduct and attitude should influence final grades of students (Cross & Frary, 1996; Frary, Cross & Weber, 1993).

One contributing factor may be that after 16 years of obtaining grades based on factors other than academic achievement, teachers-in-training have a difficult time accepting theoretical principles that do not match with their personal experience. Many beliefs about school practices are well established before students enter college and often are resistant to change (Britzman, 1986; 1991; Ginsberg & Clift, 1990; Holt-Reynolds, 1992; Pajares, 1992; Richardson, 1996). Undergraduate teacher education majors, when asked about the criteria that should be used for their own grades, believe that "effort" is more important than amount of content learned (Placier, 1995).

All grading is at some level inherently subjective. However, teachers need to recognize the subjective factors in order to reduce them as much as possible to increase the objectivity and validity of their assessment and grading practices (Guskey, 1994; Nottingham, 1988; O'Conner, 1995; Ornstein, 1994). In addition to instruction on how to assess and grade on sound principles of measurement, research suggests that preservice teachers need hands-on experience in grading students and how to work with cooperating teachers who assess and grade in ways different than those learned by the preservice teachers (Barnes, 1985; Lomax, 1996).

But it is not only preservice and inservice teachers who fail to base grading decisions on sound measurement principles. Research on grading practices of undergraduate teaching faculty indicate that faculty in applied life disciplines (such as education) tend to consider student development and personal growth factors in assessing students academic performance, while faculty in hard, pure disciplines (such as math and science) tend to focus more strictly on academic achievement (Barnes, Bull, Campbell, & Perry, 1998). Thus, it appears that teacher educators may model inappropriate grading practices in opposition to principles presented in measurement texts. Wideen, Mayer-Smith, & Moon (1998) suggest that students in teacher education programs often respond more to the "hidden" curriculum than the curriculum intentionally presented by their professors. For example, they pay more attention to the actual grading practices that their professors use than the practices that the professors teach them that they should use.

In addition, Brookhart (1998) suggests that classroom assessment and grading practices are at the center of effective management of classroom instruction and learning. Through the use of real classroom scenarios, preservice teachers need to be taught assessment strategies in relationship to instruction and not as decontextualized measurement principles. As the past

president of the American Educational Research Association, Lorrie Shepard (2000), has stated: "The transformation of assessment practices cannot be accomplished in separate tests and measurement courses, but rather should be a central concern in teaching methods courses" (p. 4).

Although assigning grades is probably the most important measurement decision that classroom teachers make, the coverage of grading in assessment textbooks is often not as fully developed as other measurement topics that are less relevant to teachers day-to-day assessment practices (Airasian, 1991; Lomax, 1996). Nor is the topic of classroom grading practices covered in much depth in many teacher education programs (Schafer, 1991; Stiggins, 1991; 1999). A recent random sample of teacher education programs indicates that less than half of teacher education programs require a measurement course (see appendix A). In those programs that do require a measurement/assessment course, often the course is focused on either "informal" assessment of typical students, or the assessment of students with special needs. In addition, fewer than half of the 50 states require specific coursework on assessment for their initial certification of teachers (Lomax, 1996; O'Sullivan & Chalnick, 1991; Stiggins, 1999).

What the literature suggests is that educators at all levels make decisions when assigning grades that are not based on sound principles of validity and reliability that ensure the grade is a meaningful vehicle for communication of a student's level of academic achievement. The literature also suggests that students in teacher education programs may be more influenced by the grading practices they have experienced as students in the past, as well as in their current courses taught by their education professors, than by what they learn about assessment and grading in their courses. Additionally, teachers in the field, as products of teacher education programs, seem to exhibit grading practices that confirm that they have not been influenced by measurement courses. This is either because they did not take any assessment courses, or

because the focus of the courses did not take into account the realities of classroom life, or their long-held beliefs about grading were left unchallenged.

Research Methods

This study investigates some of the reasons preservice and inservice teachers and teacher-educators use in making assessment and grading decisions. A survey was developed to collect data on educators' perspectives about assessment and grading issues as addressed in a case study (see Appendix B). The case study presents a realistic grading dilemma in which a teacher (Sarah) assigns grades based on criteria that could be considered both invalid and unreliable. The use of a case study was used for several reasons. First, over the past decade one of the authors of this research has used this case to help students in his educational psychology course reflect on the complexity of classroom assessment. Each semester it is the case that students are the most reluctant in accepting alternative perspectives on grading than the one that has been instilled in them over their careers as students. Specifically the perspective that effort should be included in determining a student's grade. Also, the research by Brookhart (1993) on teachers' grading practices involved collection of data through the use of a series of one-paragraph scenarios about teacher's grading practices. Although we found this method good at attempting to get teachers to reflect on and respond to "realistic" classroom situations, we thought that they lacked the true complexity of classrooms and the multiplicity of factors that teachers must take into consideration in making decisions. The case study of "Sarah Hanover" is four pages long and allows for that complexity to be developed (Silverman, Welty, & Lyon, 1996).

Research participants were asked to read and analyze the case and provide a solution to the grading dilemma along with rationale for their decisions. The major dilemma is that Sarah has given the most knowledgeable student (James) in her math class [98% test average, best

peer-tutor, active participant in class, pleasant personality] a letter grade of "B" because he never turned in homework. The parents of the student want Sarah to change the student's grade to an "A" to reflect his knowledge of math. Questions of the validity of the grade and the reliability of homework are central issues in the case. Research participants were asked to respond to the following four questions about grading and homework after reading the case:

- Q1: Should Sarah change James' grade? No Yes Why/why not?
- Q2: What letter grade does James deserve? A B C D F Why?
- Q3: What do you believe should be the purpose of homework?
- Q4: What do you believe should be the purpose of a grade?

Subjects

Data were collected from the following groups of educators:

- 153 pre-service teachers- the majority (151) of whom were enrolled in an undergraduate educational psychology survey course. These students were in their third year of the various teacher-education programs (El. Ed., Special Ed., Secondary Ed., Art Ed., etc.) in a School of Education at a private liberal arts college in upstate New York. These undergraduate students do not take a course in measurement. Their primary exposure to measurement theory and grading is as one of the major topics covered in the undergraduate educational psychology survey course (approximately 2 weeks out of a 15-week semester). They had read a chapter on classroom assessment and had 2-3 class periods of lecture/discussion on the chapter before reading the case and completing the questionnaire.
- 49 pre-service teachers enrolled in either a graduate educational psychology survey course in New York (n = 37) or graduate education courses at a California State university (n = 12). The majority (n = 35) of the New York graduate students were students not having an

undergraduate education degree. They were taking the survey course as a foundation for their other Master's level education courses and were being introduced to measurement theory and classroom grading issues for the first time (again in 2 weeks of a 15-week semester). The New York graduate students had read a chapter on classroom assessment and had 2-3 class periods of lecture/discussion on the chapter before reading the case and completing the questionnaire. The California graduate students were doing their student teaching experience that is part of the credential program. These students had taken an Educational Psychology course that includes approximately six hours of discussion on classroom assessment.

- 81 practicing high school and elementary teachers teaching in New York (n = 44) or California (n = 37). The majority (31) of the New York teachers indicated that they had taken a "graduate level course in tests and assessments." Seventeen of the 37 California graduate students were taking an assessment course when the data were collected.
- 34 school of education professors teaching in New York (n = 23) or California (n = 11). These included professors in departments of teacher education, special education, educational psychology, reading, communication disorders, and counseling.

Students in the courses were asked to complete the survey as a voluntary non-graded activity. In New York, high school and college teachers were provided a $5 incentive (gift certificate/movie pass) to voluntarily complete the survey. The response rate for these high school teachers was 63% (44 of 70 surveys sent) and for college faculty 51% (23 of 45 surveys).

Analysis and Results

In analyzing the data for Question 1 (Should Sarah change James' grade? Yes/No, why/why not?), subjects' responses could be categorized into five "yes" reasons and four "no"

reasons. Those who responded "yes" stated they would change James' grade from the assigned "B" to an "A" due to one or more of the following reasons:

1. The teacher's classroom grading policy was unclear
2. The student's classroom contributions were sufficient to earn a higher grade
3. The student's peer tutoring exhibited extra effort
4. The student could make up work or do extra assignments
5. The teacher needed to met the individual needs of the student

Those who responded "no" stated they would not change James' grade from the assigned "B" due to one or more of the following reasons:

1. The student had not met or fulfilled the teacher's stated requirements to earn an "A"
2. It would be unfair to other students to change the student's grade
3. By changing the grade it would give the wrong impression to the student
4. If the grade was change, then the teacher would have to change grades for all students

In analyzing the data for Question 2 (What letter grade does James deserve? A/B/C/D/F, Why?), subjects' responses could be categorized into three reasons for an "A", four reasons for a "B", and two reasons for an "A-/B+" grade. Those who responded that they would give James a grade of "A" provided the following reasons:

1. The teacher's grading policy was unclear and/or the student should not be penalized for the teacher's lack of clear communication
2. The student demonstrated through his classroom engagement that he was an "A" student
3. The student had demonstrated that he knew the material at an "A" level

Those who responded that they would give James a grade of "B" provided the following reasons:

1. The teacher should lower the grade by at least 10% for the homework not turned in

2. The student didn't fulfill all the requirements to earn an "A"
3. It would be unfair to other students who had completed the homework to give the "A"
4. The student needs to suffer the consequences of his not turning in homework

Those who responded that they would give James a grade of "A-/B+" provided the following reasons:

1. It was a way to compromise with the student and his parents
2. The lack of homework should only drop the 98% average to an A-/B+ overall average

The data analyzed so far related to Questions Q1 and Q2 suggest that the majority of all four groups of subjects believe a student's grade should be lowered for non-achievement related reasons (see Figure 1 and Figure 2). This violates the principle of validity of a grade by not providing an accurate statement of the student's academic achievement.

Most (86%) undergraduate students (preservice teachers) believed that James deserved a "B" and 91% indicated that Sarah should not change James grade. Only 14 of the 153 undergraduates (9%) said that the "B" grade was invalid and inappropriate and should be changed to an "A" to make it an accurate assessment of the student's achievement. The reason given most often (96%) for lowering the grade was that James failed to meet the requirements of the teacher by not turning in homework and therefore deserved a "B" (see Figure 3). In class discussions with the inservice students, many argued from the position that the grade should remain a "B" because that was the way that they had always been graded. What was important was to put forth the effort to fulfill the teacher's requirements and issues of validity seemed to have very little influence. So although the students had been exposed to the principles of validity and reliability, most failed to apply them in a simulated, but realistic situation.

Graduate students were much more likely to indicate that James deserved a higher grade than a "B" (30%), but most (86%) stated that Sarah should not change James' grade. Graduate students were more likely to relate the purpose of a grade to indicate the level of knowledge of the material by James. Among the graduate students, 15 of the 49 (31%) indicated the grade was invalid and should be changed to an "A" or "A-/B+". Most students (69%) felt that the grade should remain a "B", even though the student had a 98% average, most (93%) because he failed to put forth any effort to do the required homework.

Results also show that 88% of the practicing teachers were against changing James grade, although only 78% indicated that James deserved the "B" grade. Similar to the other three groups, a majority (93%) cited "homework as a requirement in the grading criteria used by Sarah" as the reason for not changing the grade. Only 13% cited reasons other than "homework as a required assignment."

Of all four groups surveyed, college professors had the highest percentage of respondents (32%) who thought that Sarah should change James' grade. However, the number of college professors who were against changing the grade (68%) was more than twice those in favor of changing the grade. The majority (59%) of the college professors stated that the "B" was the grade James deserved primarily (91%) because he failed to fulfill the teacher's requirements.

All four groups of respondents cited "homework as a requirement" as the major reason (all above 90%) for not changing James' grade. This would seem to imply that a grade is used to communicate not only how much content knowledge one has achieved, but also how well one has complied with the teacher's requirements.

For those respondents who did believe that James' grade should be changed to an "A" or "A-/B+", the reasons given for changing the grade varied across groups (see Figure 4). College

professors, 32% of whom said the grade should be raised, gave two primary reasons: the teacher's classroom grading policy was unclear (73%), and James' classroom contributions sufficiently indicated high academic achievement (73%). The 12% of practicing teachers who believed the grade should be raised, most often gave the reason of sufficient classroom contributions (40%), but also often gave reasons related to the unclear grading policy (30%) and the need for the teacher to meet the individual needs of James as an exceptionally talented student (20%). It is interesting to note, that it was only a few practicing teachers that mentioned the student's individual needs as a reason for changing the grade. No one in any of the other groups mentioned this. The 14% of the graduate students that said they would change the grade, most often mentioned the unclear grading policy as the reason (57%). The 9% of the undergraduate students who indicated they would change the grade most often give the reason that James' classroom contributions were sufficient (64%).

[Analysis regarding Question 3 (What do you believe should be the purpose of homework?) and Question 4 (What do you believe should be the purpose of a grade?) is not yet completed.]

Conclusions and Educational Implications

Concerns about the validity and reliability of grades for communicating meaningful information about students' academic progress have been raised for a long time (e.g., see Starch & Elliot, 1912; 1913a; 1913b; Adam, 1932; Johnson, 1935). In addition, trying to help teachers to understand the purpose and effective functions of grades in the overall evaluation system has been addressed repeated in the literature and textbooks devoted to classroom assessment of students (e.g., Airasian, 2000; Brookhart, 1993; Cross & Frary, 1996; Gredler, 1999; Guskey, 1996; Linn & Gronlund, 2000; Lysne, 1984; Marzano, 2000; O'Connor, 1995; Stiggins, 2001).

However, there seems to be little progress being made based on the findings of the research presented here.

This research, in conjunction with previous research in this area, suggests two major thrusts that need to occur in reforming grading practices. First, in those situations in which a teacher must summarize and communicate a student's classroom progress in an academic subject through a single report card grade, there must be a consensus that the grade represents the most accurate statement of the student's academic achievement, and only academic achievement. This is the essence of valid assessment. To include non-academic criteria, such as the student's effort, compliance, attitude, or behavior, makes the grade impossible to interpret in any meaningful way.

In order for teachers to act consistently in assigning valid grades based only on achievement criteria, a second major initiative needs to be undertaken to help teachers make good grading decisions based on fundamental measurement principles. This initiative is best addressed through teacher education programs and entails several dimensions.

In addressing this problem as it relates to preparing preservice undergraduate and graduate students, there are at least four areas that teacher education programs need to address. First, students' long held beliefs about the purpose and use of grades need to be challenged by teacher educators. Students' beliefs and value systems related to grades need to be exposed and examined to help them understand the unscientific basis of their grading beliefs. Second, once these beliefs are exposed, instructors must provide students with the theoretical base for good assessment and grading practices as explicated by measurement experts which would replace students' naive notions of assessment and grading. This could be either though self-contained measurement courses taught in a relevant manner by educational psychologists, or integrated into

methods courses through collaboration between educational psychology and teacher education departments. It would help if more teacher education programs required assessment courses and more attention to effective and meaningful grading practices were addressed in measurement textbooks. Third, teacher education students need to be provided with opportunities to encounter grading activities before they are placed into student teaching in order to practice applying assessment principles and theory to classroom grading issues. One way to accomplish this is through the use of case studies that focus on assessment and grading dilemmas often faced by real teachers. Finally, during student teaching experiences, education majors must be given the opportunity in conjunction with their cooperating teachers and the support of their college supervisors, to actually develop and implement an evaluation and grading plan.

The above suggestions have additional implications for teacher education programs as related to the research findings presented in this paper and found in related research literature. Schools of education need to work with school district teachers to help improve the communication system for which grades function as the vehicle. Providing in-service "assessment & grading" workshops for practicing teachers, especially those operating as cooperating teachers, might help to establish a consensus of what criteria should be used for determining and assigning grades.

However, the area that may be the most difficult to address is the change in the grading practices that we as college professors use in evaluating our students. As long as preservice and inservice teachers take classes from education professors who base grading decisions on more than academic achievement, students will have little reason to believe what we say and practice what we preach about assessment and grading. As educational psychologists, we need to model sound grading practices in our own courses, as well as inform our school of education colleagues

in other departments of the validity issues involved in accurate communication of student achievement.

In summary, we believe that this research could be used to begin to break the unsound "culture of grading" practices that prevail throughout the education system. The value of this project for professional development is that it provides information to teacher-educators about the need to improve the coverage of the topic of assessment and grading in educational psychology and teacher education courses. Perhaps more importantly, the results of this study could help current and future teachers (both K-12 and Higher Education) to consider the ways they evaluate and grade students so that more valid and reliable measures of academic achievement of their students are accurately communicated to others.

References

Adams, W. L. (1932). Why teachers say they fail pupils. *Educational Administration and Supervision, 18*, 594-600.

Airasian, P. W. (2000). *Assessment in the Classroom: A Concise Approach, 2nd ed.*, Boston: McGraw Hill.

Airasian, P. W. (1991). Perspectives on measurement instruction. *Educational Measurement: Issues and Practice, 10* (1), 13-16, 26.

Bailey, J, & McTighe, J. (1996). Reporting achievement at the secondary level: What and how. In T. R. Guskey (Ed.), *ASCD Yearbook -1996: Communicating Student Learning* (pp. 119-140). Alexandria, VA: Association for Supervision and Curriculum Development.

Barnes, L. B., Bull, K. S., Campbell, N. J., & Perry, K. M. (1998). *Discipline-related differences in teaching and grading philosophies among undergraduate teaching faculty*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, April, 1998.

Barnes, S. (1985). A study of classroom pupil evaluation: The missing link in teacher education. *Journal of Teacher Education, 36* (4), 46-49.

Britzman, D. P. (1991). *Practice makes practice*. New York: State University of New York Press.

Britzman, D. P. (1986). Cultural myths in the making of a teacher: Biography and social structure in teacher education. *Harvard Educational Review, 56*(4), 442-456.

Brookhart, S. M. (1998). *Teaching about grading and communicating assessment results*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Diego, April, 1998.

Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in*

Education, 7 (4), 279-301.

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement, 30* (2), 123-142.

Cizek, G. J. (1996). Grades: The final frontier in assessment reform. *NASSP Bulletin, 80* (584), 103-110.

Cizek, G. J. et al. (1995). *Further investigation of teachers' assessment practices*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, April, 1995.

Cross, L. H. & Frary, R. B. (1996). *Hodgepodge grading: endorsed by students and teachers alike*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York, April, 1996.

Frary, R. B., Cross, L. H., & Weber, L. J. (1993). Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instruction in measurement. *Educational Measurement: Issues and Practice, 12*(3), 23-30.

Friedman, S. J. & Frisbie, D. A. (2000). Making report cards measure up. *The Education Digest, 65* (5), 45-50.

Friedman, S. J. & Frisbie, D. A. (1995). The influence of report cards on the validity of grades reported to parents. *Educational and Psychological Measurement, 55* (1), 5-26.

Gallagher, J. D. (1998). *Classroom Assessment for Teachers*. Upper Saddle River, NJ: Merrill/Prentice Hall.

Ginsburg, M. B. & Clift, R. T. (1990). The hidden curriculum of preservice teacher education. In W. R. Houston (Ed.), *Handbook of research on teacher education* (pp. 450-465). New York: Macmillan.

Gredler, M.E. (1999). *Classroom Assessment and Learning*. New York: Longman.

Guskey, T. R. (Ed.). (1996). *ASCD Yearbook -1996: Communicating Student Learning*. Alexandria, VA: Association for Supervision and Curriculum Development.

Guskey, T. R. (1994). Making the grade: What benefits students? *Educational Leadership, 52* (2), 14-20.

Holt-Reynolds, D. (1992). Personal history-based beliefs as relevant prior knowledge in coursework: Can we practice what we preach? *American Educational Research Journal, 29*(2), 325-349.

Johnsen, S. (1995). How did you do? The myths and facts about grades. *Gifted Child Today Magazine, 18* (2), 10-13.

Johnson, F. W. (1935). *The Administration and Supervision of the High School*. Boston: Ginn & Co.

Lake, K. & Kafka, K. (1996). Reporting methods in grades K-8. In T. R. Guskey (Ed.), *ASCD Yearbook -1996: Communicating Student Learning* (pp. 90-118). Alexandria, VA: Association for Supervision and Curriculum Development.

Linn, R. L. & Gronlund, N.E. (2000). *Measurement and Assessment in Teaching, 8th ed.*, Englewood Cliffs, NJ: Merrill/Prentice Hall.

Lomax, R. G. (1996). On becoming assessment literate: An initial look at preservice teachers' beliefs and practices. *Teacher Educator, 31* (4), 292-303.

Lysne, A. (1984). Grading of student's attainment: Purposes and functions. *Scandinavian Journal of Educational Research, 28*(3), 149-165.

Marzano, R. J. (2000). *Transforming Classroom Grading*. Alexandria, VA: Association for Supervision and Curriculum Development

National Education Longitudinal Study (1988). *What do student grades mean?*

Differences across schools. Education Research Report.

Nitko, A. J. (2001). *Educational Assessment of Students, 3rd ed.*, Upper Saddle River, NJ: Merrill/Prentice Hall.

Nottingham. M. (1988). Grading practices--Watching out for land mines. *NASSP Bulletin, 72* (507), 24-28.

O'Conner, K. (1995). Guidelines for grading that support learning and student success. *NASSP Bulletin, 79* (571), 91-101.

Olson, G. H. (1989). *On the validity of performance grades: The relationship between teacher-assigned grades and standard measures of subject matter acquisition.* Paper presented at the Annual Meeting of the national Council on Measurement in Education, San Francisco, March, 1989.

Oosterhof, A. (2001). *Classroom Applications of Educational Measurement, 3rd ed.*, Upper Saddle River, NJ: Merrill/Prentice Hall.

Ormrod, J. E. (2000). *Educational Psychology: Developing Learners, 3rd ed.*, Upper Saddle River, NJ: Merrill/Prentice Hall.

Ornstein, A. C. (1994). Grading practices and policies: An overview and some suggestions. *NASSP Bulletin, 78* (561), 55-64.

O'Sullivan, R. G. & Chalnick, M. K. (1991). Measurement-related course work requirements for teacher certification and recertification. *Educational Measurement: Issues and Practice, 10* (1), 17-19, 23.

Placier, M. (1995). "But I have to have an A": Probing the cultural meanings and ethical dilemmas of grades in teacher education. *Teacher Education Quarterly, 22* (1), 45-63.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct.

Review of Educational Research, 62 (3), 307-332.

Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula, T. Buttery, & E. Guyton (Eds.), *Handbook of research on teacher education* (2nd ed., pp. 102-119). New York: Macmillan.

Schafer, W. D. (1991). Essential assessment skills in professional education of teachers. *Educational Measurement: Issues and Practice, 10* (1), 3-6,12.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29* (7), 4-14.

Silverman, R, Welty, W. & Lyon, S. (1996). *Case studies for teacher problem solving, 2nd ed.*, New York: McGraw-Hill.

Snowman, J., Biehler, R. F., & Bonk, C. J. (2000). *Psychology Applied to Teaching, 9th ed.*, Boston: Houghton Mifflin.

Starch, D. & Elliot, E. C. (1912). Reliability of grading of high-school work in English. *The School Review, 20*, 442-457.

Starch, D. & Elliot, E. C. (1913a). Reliability of grading work in mathematics. *The School Review, 21*, 254-259.

Starch, D. & Elliot, E. C. (1913b). Reliability of grading work in history. *The School Review, 20*, 676-681.

Stiggins, R. J. (2001). *Student-Involved Classroom Assessment, 3rd ed.*, Upper Saddle River, NJ: Merrill/Prentice Hall.

Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice, 18*(1), 23-27.

Stiggins, R. J. (1991). Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice, 10* (1), 7-12.

Stiggins, R. J. (1988). Make sure your teachers understand student assessment. *Executive Educator, 10* (8), 24-26.

Stumpo, V. M. (1997). 3-tier grading sharpens student assessment. *The Education Digest, 63* (4), 51-54.

Thorndike, R. M. (1997). *Measurement and Evaluation, 6th ed.* Upper Saddle River, NJ: Merrill/Prentice Hall.

Tombari, M. & Borich, G. (1999). *Authentic Assessment in the Classroom*. Upper Saddle River, NJ: Merrill/Prentice Hall.

Wideen, M., Mayer-Smith, J. & Moon, B. (1998). A critical analysis of the research on learning to teach: Making the case for an ecological perspective on inquiry. *Review of Educational Research, 68* (2), 130-178.

Appendix A

Percentage of undergraduate and graduate education programs requiring/not requiring a measurement/assessment course at a random sample (n) of colleges and universities

	N*	n	# of programs	No	Yes	Yes	Yes
US Colleges	1489	16	47	53%	32%	4%	11%
CA Colleges	119	13	26	73%	15%	7%	4%
NY Colleges	135	15	45	67%	20%	0%	13%

No = No measurement/assessment course required
Yes = General Measurement/assessment course required
Yes = Measurement/assessment course required - focus on informal assessment
Yes = Measurement/assessment course required - focus on students with special needs

*SOURCE: http://www.utexas.edu/world/univ/state/
Web U.S. Universities, by State
"This page of UT Austin Web Central contains a list of regionally-accredited U.S. universities organized by state. ... This list includes only 4-year institutions."

Appendix B

Directions

This is an anonymous survey. DO NOT place your name on it. Please keep all of the materials attached. On this first section of the survey please complete all the following items that are relevant to you and your background. **Circle** all the information that best describes you.

- **Age:** 11-14 15-18 19-22 22-26 27-30 31-34 35-39 40-43 44-47 48-51 52+
- **Sex:** Female Male
- **Current grade level (K-12):** K-6th student Middle/Junior H. S. student H. S. student
  K-6th teacher Middle/Junior H.S. teacher H.S. teacher
  Reading teacher Special Ed teacher Administrator
- **Current school level (college):** Undergraduate Student Graduate Student Faculty
- **I have been a K-12 teacher for:** 1-3 years 4-6 yr. 7-9 yr. 10+ yr.
- **I have been a K-12 administrator for:** 1-3 years 4-6 yr. 7-9 yr. 10+ yr.
- **I have been a college teacher for:** 1-3 years 4-6 yr. 7-9 yr. 10+ yr.
- **College Major/Department to which you are associated:**

  Music Education Art Education Early Childhood Ed
  Elementary Education Special Ed/Elem. Ed (SEED) Special Education
  English/Secondary Ed Math/Secondary Ed Biology/Secondary Ed
  Social Studies/Secondary Ed Spanish/Secondary Ed Chemistry/Secondary Ed
  Educational Psychology School Psychology Communication Disorders (CMD)
  Ed Administration Reading Counseling Applied Technology Ed (ATE)
- **I have taken a Graduate level course on educational tests and measurement:** No Yes
  - **I have taken EPY529: Mental & Educational Measurement at St. Rose:** No Yes
- **I have taken an Undergraduate level course on educational tests and measurement:** No Yes

Directions

Please carefully read the attached case study, *Sarah Hanover*[1]. After reading it, answer the questions on the following pages as concisely as possible. If you find it necessary, you may attach additional comments.

---

[1] Silverman, R., Welty, W. & Lyon, S. (1996). *Case studies for teacher problem solving, 2nd ed.* New York: McGraw-Hill.

Please answer the following questions based on the assumption that Sarah must record one of the following letter grades for James on his report card/school record [A, B, C, D, F].

- Should Sarah change James' grade? No Yes
- Why/Why not? _____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

- What letter grade does James deserve? A B C D F
- Why? _____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

- What do you believe should be the purpose of homework?

- What do you believe should be the purpose of a grade?

Thank you for your time and thoughtful completion of this survey.

Dr. James D. Allen

Figure 1. Should Sarah change James' grade?

	Yes	No
Undergraduate	9%	91%
Grad	14%	86%
Prac Teachers	12%	88%
College Prof	32%	68%

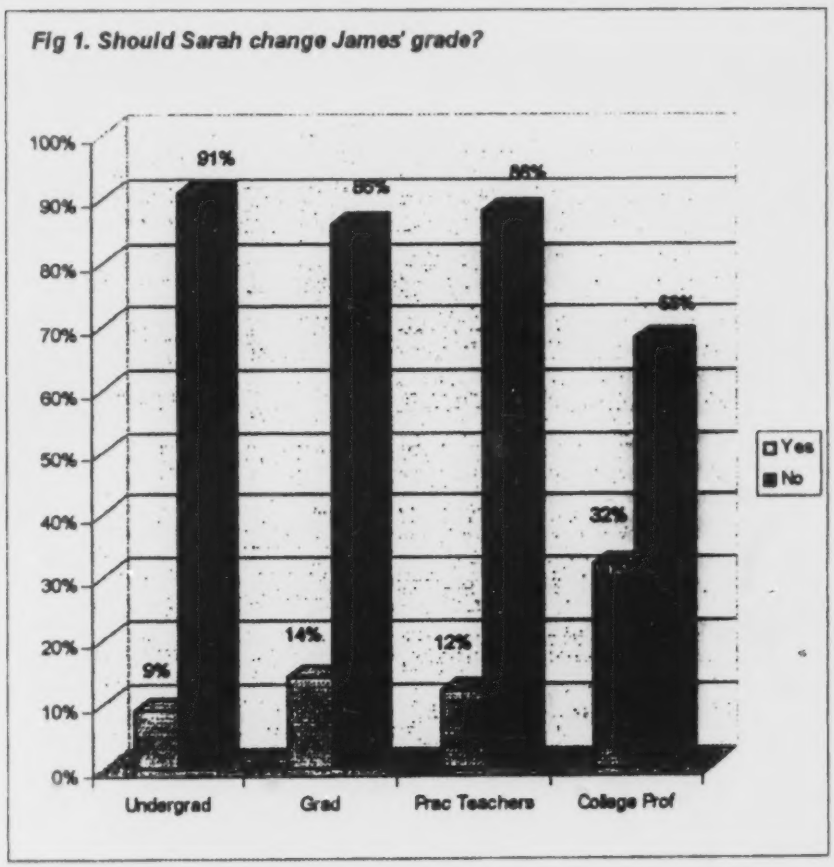
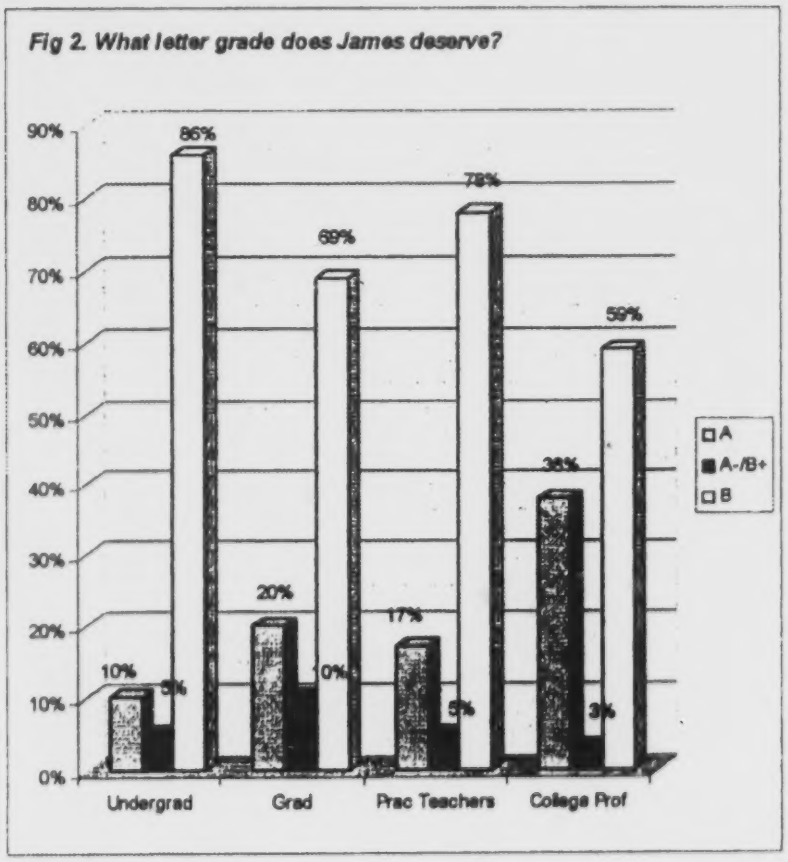


Figure 2. What letter grade does James deserve?

	A	A-/B+	B
Undergraduate	10%	5%	86%
Grad	20%	10%	69%
Prac Teachers	17%	5%	78%
College Prof	38%	3%	59%



BEST COPY AVAILABLE

Figure 3. Reasons for not changing the grade

	Requirement	Fairness	Impression	Other Students
Undergraduate	96%	7%	4%	3%
Graduate	93%	5%	0%	5%
Prac Teachers	93%	3%	4%	6%
College Prof	91%	9%	9%	0%

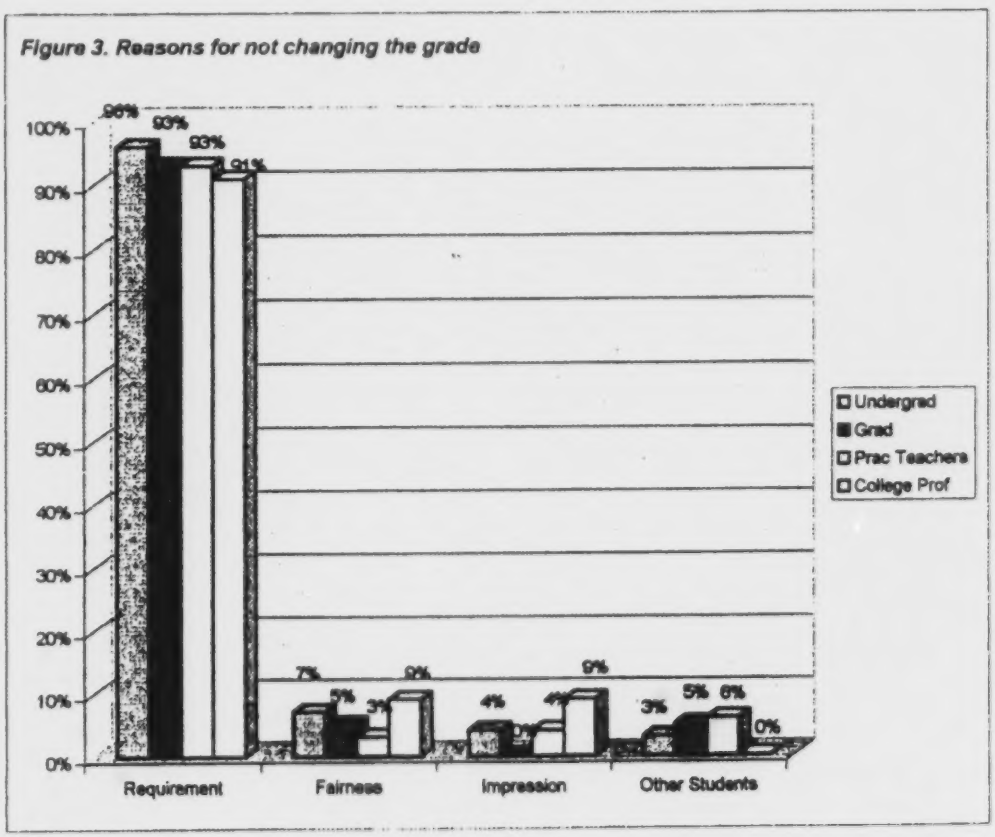
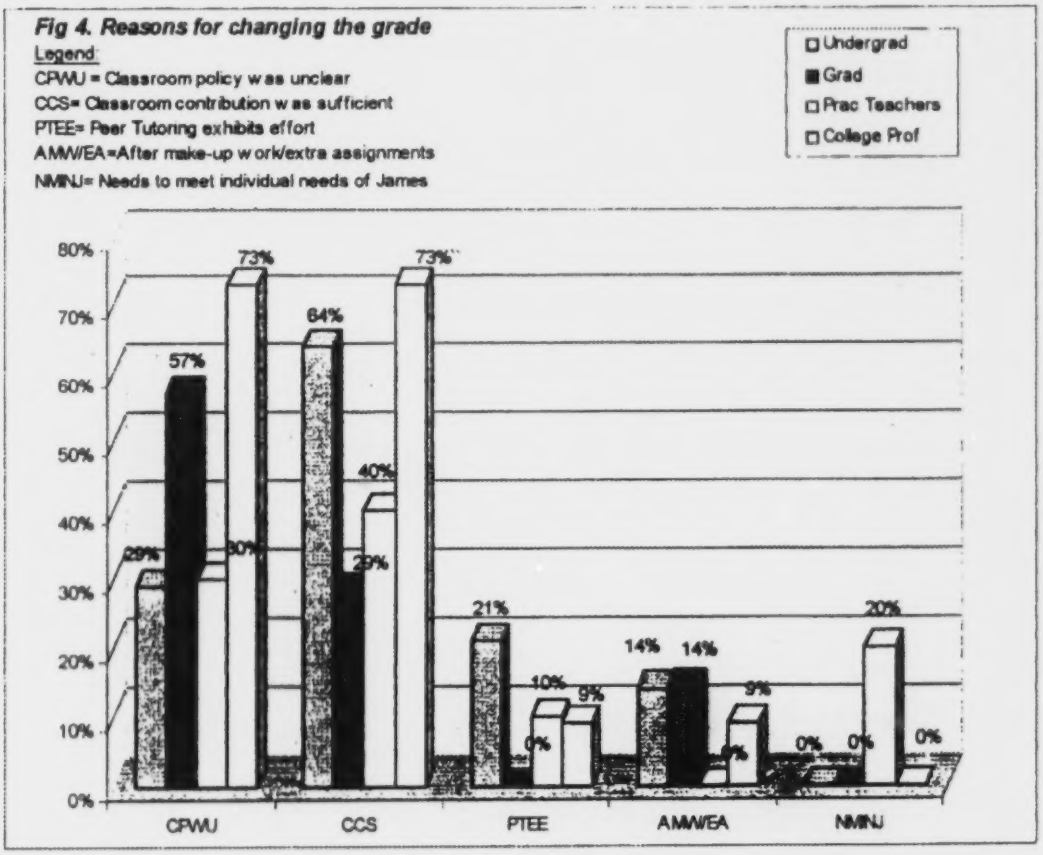


Figure 4. Reasons for changing the grade

	CPWU	CCS	PTEE	AMW/EA	NMINJ
Undergraduate	29%	64%	21%	14%	0%
Graduate	57%	29%	0%	14%	0%
Prac Teachers	30%	40%	10%	0%	20%
College Prof	73%	73%	9%	9%	0%



BEST COPY AVAILABLE